

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-259083

(43) 公開日 平成11年(1999) 9月24日

(51) Int.Cl.⁶G 1 0 L 3/00
5/02

識別記号

F I

G 1 0 L 3/00
5/02H
J

審査請求 未請求 請求項の数15 O L (全 11 頁)

(21) 出願番号 特願平10-57248

(22) 出願日 平成10年(1998) 3月9日

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 奥谷 泰夫

東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

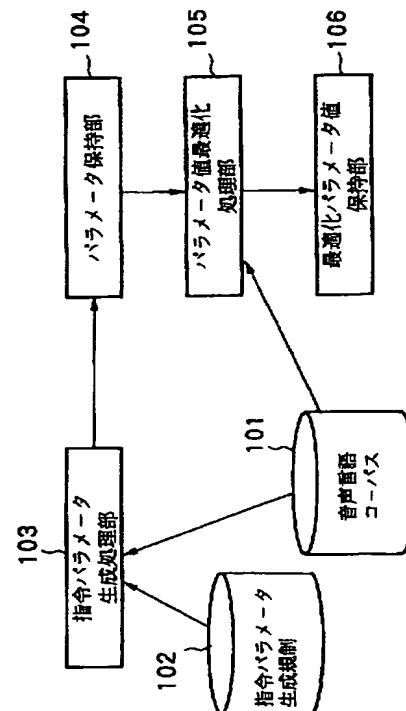
(74) 代理人 弁理士 大塚 康德 (外2名)

(54) 【発明の名称】 音声合成装置および方法

(57) 【要約】

【課題】 話者に応じた基本周波数パターンを生成可能とし、自然な合成音声を得ることを可能とする。

【解決手段】 音声言語コーパス101には、基本周波数パターンを含む音声情報が格納される。指令パラメータ生成処理部103は、音声言語コーパス101の音声情報に指令パラメータ生成規則102を適用して基本周波数モデルの指令パラメータを生成する。パラメータ値最適化処理部105は生成された指令パラメータに対応するパラメータ値を、音声言語コーパス101に格納された基本周波数パターンに基づいて最適化する。最適化されたパラメータ値は、最適化パラメータ値保持部106によって保持され、音声合成処理における基本周波数パターン生成時に使用される。



【特許請求の範囲】

【請求項1】 基本周波数ボタンを含む音声情報を格納した格納手段と、

前記音声情報に所定の指令パラメータ生成規則を適用して基本周波数モデルの指令パラメータを生成する生成手段と、

前記生成手段で生成された指令パラメータに対応するパラメータ値を、前記格納手段に格納された基本周波数ボタンに基づいて最適化する学習手段と、

前記学習手段で得られたパラメータ値を保持する保持手段と、

前記保持手段に保持されたパラメータ値を用いて基本周波数ボタンを生成し、音声合成を行う合成手段とを備えることを特徴とする音声合成装置。

【請求項2】 前記格納手段に格納されている音声情報は話者別に用意されており、

前記学習手段は、話者別にパラメータ値の最適化を行い、

前記保持手段は、話者別に最適化されたパラメータ値を保持することを特徴とする請求項1に記載の音声合成装置。

【請求項3】 前記音声情報は、複数の文に関する音韻系列および時間長、アクセント型、モーラ数、ポーズ位置、ポーズ長、係り受け関係、基本周波数ボタンを含むことを特徴とする請求項1に記載の音声合成装置。

【請求項4】 前記合成手段は、文の言語解析結果に前記所定の指令パラメータ生成規則を適用して指令パラメータを生成する手段と、

前記指令パラメータに対応するパラメータ値を前記保持手段より獲得する手段と、

獲得されたパラメータ値に基づいて基本周波数ボタンを生成する手段と、

生成された基本周波数ボタンを用いて音声合成処理を行う手段とを備えることを特徴とする請求項1に記載の音声合成装置。

【請求項5】 前記学習手段は、前記指令パラメータに候補パラメータ値を代入して基本周波数ボタンを生成する第1手段と、

前記第1手段で生成された基本周波数ボタンと前記格納手段に用意された基本周波数ボタンとの差が小さくなるようにパラメータ値を設定する第2手段とを備えることを特徴とする請求項1に記載の音声合成装置。

【請求項6】 前記第1手段及び前記合成手段による基本周波数ボタンの生成において、フレーズ指令の有効範囲内で平坦部の発生を解消するための補正を行うことを特徴とする請求項5に記載の音声合成装置。

【請求項7】 前記補正は、前記フレーズ指令の有効範囲に負の傾きを有する直線を重畳することで行うことを特徴とする請求項6に記載の音声合成装置。

【請求項8】 メモリに格納された基本周波数ボタンを

含む音声情報に所定の指令パラメータ生成規則を適用して基本周波数モデルの指令パラメータを生成する生成工程と、

前記生成工程で生成された指令パラメータに対応するパラメータ値を、前記メモリに格納されている基本周波数ボタンに基づいて最適化する学習工程と、

前記学習工程で得られたパラメータ値を保持する保持工程と、

前記保持工程で保持されたパラメータ値を用いて基本周波数ボタンを生成し、音声合成を行う合成工程とを備えることを特徴とする音声合成方法。

【請求項9】 前記メモリに格納されている音声情報は話者別に用意されており、

前記学習工程は、話者別にパラメータ値の最適化を行い、

前記保持工程は、話者別に最適化されたパラメータ値を保持することを特徴とする請求項8に記載の音声合成方法。

【請求項10】 前記音声情報は、複数の文に関する音韻系列および時間長、アクセント型、モーラ数、ポーズ位置、ポーズ長、係り受け関係、基本周波数ボタンを含むことを特徴とする請求項8に記載の音声合成方法。

【請求項11】 前記合成工程は、文の言語解析結果に前記所定の指令パラメータ生成規則を適用して指令パラメータを生成する工程と、

前記指令パラメータに対応するパラメータ値を前記保持工程で保持されたパラメータ値より獲得する工程と、

獲得されたパラメータ値に基づいて基本周波数ボタンを生成する工程と、

生成された基本周波数ボタンを用いて音声合成処理を行う工程とを備えることを特徴とする請求項8に記載の音声合成方法。

【請求項12】 前記学習工程は、前記指令パラメータに候補パラメータ値を代入して基本周波数ボタンを生成する第1工程と、

前記第1工程で生成された基本周波数ボタンと前記メモリに用意された基本周波数ボタンとの差が小さくなるようにパラメータ値を設定する第2工程とを備えることを特徴とする請求項8に記載の音声合成方法。

【請求項13】 前記第1工程及び前記合成工程による基本周波数ボタンの生成において、フレーズ指令の有効範囲内で平坦部の発生を解消するための補正を行うことを特徴とする請求項12に記載の音声合成方法。

【請求項14】 前記補正は、前記フレーズ指令の有効範囲に負の傾きを有する直線を重畳することで行うことを特徴とする請求項13に記載の音声合成方法。

【請求項15】 音声合成処理を行うための制御プログラムを格納する記憶媒体であって、該制御プログラムが、

メモリに格納された基本周波数ボタンを含む音声情報に

所定の指令パラメータ生成規則を適用して基本周波数モデルの指令パラメータを生成する生成工程のコードと、前記生成工程で生成された指令パラメータに対応するパラメータ値を、前記メモリに格納されている基本周波数ボタンに基づいて最適化する学習工程のコードと、前記学習工程で得られたパラメータ値を保持する保持工程のコードと、

前記保持工程で保持されたパラメータ値を用いて基本周波数ボタンを生成し、音声合成を行う合成工程のコードとを備えることを特徴とする記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、基本周波数制御が可能な音声合成装置及び方法に関する。

【0002】

【従来の技術】一般に、音声合成における基本周波数制御では、「日本語単語アクセントの基本周波数ボタンとその生成機構のモデル」（藤崎博也、須藤寛、日本音響学会誌、27巻、9号、pp.445～453、1971年）において提案された基本周波数生成過程モデルが用いられることが多い。この基本周波数生成過程モデルは、基本周波数の対数値の時間変化を文頭から文末に向かう緩やかな下降のフレーズ成分と、局所的な起伏のアクセント成分との和で表わす重畳型モデルである。

【0003】このモデルを使って音声合成の基本周波数制御を行なうためには、フレーズ成分の立ち上がり時点とその大きさ、アクセント成分の立ち上がり時点と立ち下がり時点とそれらの大きさなどのパラメータや、フレーズ指令およびアクセント指令の固有角周波数などのパラメータを決定する必要がある。そのための規則やパラメータの値については、例えば「日本語文章音声の合成のための韻律規則」（河合恒、広瀬啓吉、藤崎博也、日本音響学会誌、50巻、6号、pp.433～442、1994年）において提案されている。

【0004】この提案では、フレーズ指令やアクセント指令の位置や固有角周波数は、話者や文によらず一定値として用いることが一般的になっているため、話者の特徴を十分に反映できていないと言えない。

【0005】また、パラメータの最適化を行なう方法が「統計的手法による基本周波数ボタンの制御」（平井俊男ほか、日本音響学会講演論文集、2-8-3、pp.225～226、1993年）で提案されている。この方法では、音声言語コーパスに対してパラメータの最適化を行なうので、比較的安定したパラメータの値を求めることが可能である。

【0006】

【発明が解決しようとする課題】しかしながら、この方法でも、前述したようにフレーズ指令やアクセント指令の固有角周波数や指令時点を固定しているため、話者による特徴を基本周波数ボタンの曲線中に表現することが

困難である。

【0007】また、独自の指令パラメータ生成規則（実施形態に記述）を作成した場合に、この基本周波数生成過程モデルでは、文末や句末の基本周波数がほとんど一定の値、つまり、平坦になってしまうことがあり、合成音声を生じた場合に不自然さが残ってしまう原因となる。また、フレーズ成分は、ポーズが来るまでリセットされないため、フレーズ成分の重量が頻繁に起きると基本周波数が尻上がりに高くなり、不自然な合成音声になってしまう原因となる。

【0008】本発明は上述した問題に鑑みてなされたものであり、話者に応じた基本周波数ボタンを生成可能とし、自然な合成音声を得ることが可能な音声合成装置および方法を提供することを目的とする。

【0009】又、本発明の他の目的は、基本周波数ボタンの生成モデルに改良を加え、より自然な音声合成を可能とすることにある。

【0010】

【課題を解決するための手段】上記の目的を達成するための、本発明の一態様による音声合成装置は、例えば以下の構成を備える。すなわち、基本周波数ボタンを含む音声情報を格納した格納手段と、前記音声情報に所定の指令パラメータ生成規則を適用して基本周波数モデルの指令パラメータを生成する生成手段と、前記生成手段で生成された指令パラメータに対応するパラメータ値を、前記格納手段に格納された基本周波数ボタンに基づいて最適化する学習手段と、前記学習手段で得られたパラメータ値を保持する保持手段と、前記保持手段に保持されたパラメータ値を用いて基本周波数ボタンを生成し、音声合成を行う合成手段とを備える。

【0011】又、本発明によれば、上記音声合成装置に対応した音声合成方法、及び該音声合成方法をコンピュータによって実現するための制御プログラムを格納した記憶媒体が提供される。

【0012】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を詳細に説明する。

【0013】図1は、本発明の一実施形態に係る音声合成装置のオフライン処理における機能構成を示すブロック図である。図1において、101は大量の文に対する音韻系列および時間長、アクセント型、モーラ型、ポーズ位置、ポーズ長、係り受け関係、基本周波数ボタン等の音声合成情報を格納した音声言語コーパスである。102は指令パラメータを生成するための規則を格納した指令パラメータ生成規則である。103は指令パラメータ生成規則102に従って、改良型基本周波数生成モデルの各指令パラメータを、音声言語コーパス101に格納された各文について生成する指令パラメータ生成処理部である。104は指令パラメータ生成処理部103で生成したパラメータを保持するパラメータ保持部であ

る。105は音声言語コーパス101の大量の基本周波数パターンを学習サンプルとして、パラメータ保持部104で保持している各パラメータの値の最適化を行なうパラメータ値最適化処理部である。106は最適化した各パラメータの値を保持する最適化パラメータ値保持部である。なお、音声言語コーパスには、複数話者について、話者毎に音声合成情報が格納されており、本オフライン処理では、話者毎に各パラメータ値の最適化を行う。従って、最適化パラメータ保持部106は、話者毎に最適化パラメータを保持し、話者別パラメータ値辞書114（図2）を生成することになる。

【0014】図2は、本発明の一実施形態に係る音声合成装置のオンライン処理における機能構成を示すブロック図である。111は指令パラメータ生成規則102と同等の指令パラメータ生成規則である。112は指令パラメータ生成処理部103と同じく指令パラメータを生成する指令パラメータ生成処理部である。113は生成した指令パラメータを保持するパラメータ保持部である。114はオフライン処理により求めた話者別のパラメータ値を格納した話者別パラメータ値辞書である。115は話者別パラメータ値辞書114で保持しているパラメータ値と、パラメータ保持部113にて保持しているパラメータ系列とから基本周波数パターンを生成する基本周波数パターン生成処理部である。116は生成した基本周波数パターンを格納する基本周波数パターン保持部である。

【0015】本実施形態では、まずオフライン処理によって基本周波数パターンモデルの各パラメータの最適値を計算し、オンライン処理では、その結果を用いて基本周波数パターンの生成を行なう。

【0016】図3は本実施形態による音声合成装置の構成を示すブロック図である。図3において、21は制御メモリであり、図4のフローチャートに示すような制御手順に従った処理を実現するための制御プログラムを記憶する。本実施形態では、制御メモリ21をROMで構成する。22は制御メモリ21に保持されている制御手順に従って判断・演算などを行なう中央処理装置である。23はメモリであり本例ではRAMで構成される。メモリ23は、オフライン処理では、パラメータ保持部104、最適化パラメータ値保持部106におけるデータ保持領域を提供する。また、オンライン処理では、パラメータ保持部113、基本周波数パターン保持部116によるデータ保持領域を提供する。

【0017】24はディスク装置であり、オフライン処理における音声言語コーパス101、指令パラメータ生成規則102を保持する。また、オンライン処理における指令パラメータ生成規則111、話者別パラメータ値辞書114を保持する。25はバスであり、上記の各構成を相互に通信可能に接続する。

【0018】次に、オフライン処理における動作を説明

する。図4は本実施形態の音声合成装置によるオフライン処理の動作を説明するフローチャートである。まず、ステップS301では、音声言語コーパス101に含まれるすべての文について、指令パラメータ生成規則に従って指令パラメータを生成する。指令パラメータ生成規則の一例を以下に示す。

【0019】 [指令パラメータ生成規則]

(1) 入力データ

以下に示す各アクセント句ごとの情報を入力とする。

- ・モーラ数：アクセント句のモーラ数
- ・アクセント型：アクセント句のアクセント型
- ・係り先：アクセント句の係り先
- ・係り先までの距離：アクセント句の係り先のアクセント句番号－当該アクセント句番号
- ・修飾タイプ：係り先のアクセント句が同一の句（文節）の場合は、“句内”とする。文末のアクセント句の場合は、“なし”とする。他に“連体修飾”、“連用修飾”、“並列”がある。
- ・ポーズ句中でのアクセント句位置：ポーズ句の先頭のアクセント句から数えた場合のアクセント句の位置番号
- ・ポーズ句中でのモーラ位置：ポーズ句の先頭のモーラを1とした場合の当該アクセント句の先頭モーラの位置番号
- ・ポーズの時間長：アクセント句の末尾にポーズが存在する場合はその長さを、存在しない場合は0を指定する。
- ・音韻系列：アクセント句を構成する音韻の系列
- ・音韻の時間長：アクセント句を構成する各音韻の時間長。

【0020】 (2) 出力データ

ポーズ句を構成するアクセント句の情報を入力し、それらのアクセント句に対して以下のパラメータを出力する。

- ・フレーズ指令の大きさ：大きさ0以外に、4種類のフレーズ指令の大きさがある、
 - p h 1…通常の場合（小）
 - p h 2…通常の場合（大）
 - p h 3…重畳する場合
 - p h 4…ルールA-7の場合（後述）
- ・フレーズ成分のリセット：ポーズ句先頭からのフレーズ成分の重畳分をリセットするかどうかを表わす
- ・アクセント指令の大きさ：平板型は常に同じ大きさとする。有核型の場合は、2段階に規格化したアクセント指令の大きさがある、
 - a c c d 1…小
 - a c c d 2…大
- ・フレーズ成分の固有角周波数の大きさ：各フレーズ指令の大きさによって異なる値を設定する。フレーズ成分の概形を変化させるためのパラメータである、
 - a 1…通常のフレーズ指令（p h 1, p h 2）の場合

$\alpha 2$ …フレーズ指令を重畳する場合

$\alpha 3$ …ルールA-7の場合(後述)

・補正直線傾き定数: 句末、文末の傾斜のための補正曲線の傾きを決定するための定数である、

k 1…フレーズ指令の有効範囲が1000msec以下の場合

k 2…フレーズ指令の有効範囲が1000msecより大きい場合。

【0021】(3)パラメータの生成規則

先に述べた入力データを利用してパラメータ(上記出力データ)を生成するための規則について説明する。

【0022】指令パラメータを生成するための規則は3つに大別できる。ルールAとルールBは、主にフレーズ指令、アクセント指令の大きさの初期値の設定、ポーズ句の途中のアクセント句にフレーズ指令を挿入するか否か、また、挿入する場合は、累積しているフレーズ成分をリセットするか否かを決定するためのルール群である。ルールCは、フレーズ指令、アクセント指令の大きさを決定するためのルール群である。なお、ルールの適用は、ルールA、ルールB、ルールCの順に行なう。

【0023】(3-1)ルールA

ルールAを適用する場合の処理の流れは、以下の通りである。

- ・ステップ0: 文頭のアクセント句に着目する。
- ・ステップ1: 着目しているアクセント句に対し、以下に記述するルールA-1からルールA-7の条件を満足するルールを適用する。条件判定はルールの記述順に行なう。
- ・ステップ2: 後接するアクセント句が存在する場合は、後接するアクセント句に着目し、ステップ1に戻る。また、後接するアクセント句が存在しない場合は終了する。

【0024】以下にルールAに含まれる各ルール(ルールA-1~ルールA-7)について説明する。

【0025】<ルールA-1>着目しているアクセント句がポーズ句の先頭のアクセント句の場合は、フレーズ指令の大きさとしてp h 3を設定する。また、固有角周波数を $\alpha 1$ とする。また、それ以外の場合は、フレーズ指令の大きさを0とする(つまり、当該アクセント句において、フレーズ指令が存在しない)。

【0026】<ルールA-2>有核型アクセント指令の大きさとしてa c c d 1を設定する。

【0027】<ルールA-3>着目しているアクセント句が、ポーズ句の先頭でなく、9モーラより長いアクセント句の場合であって、前接するアクセント句のモーラ数が4以上の場合、大きさp h 3のフレーズ指令を挿入し、固有角周波数を $\alpha 2$ とする。それ以外の場合は、累積しているフレーズ成分を一旦リセットし、大きさp h 2のフレーズ指令を挿入し、固有角周波数は $\alpha 1$ とする。

【0028】<ルールA-4>前接するアクセント句が平板型の場合は、大きさp h 3のフレーズ指令を挿入し、固有角周波数は $\alpha 2$ とする。ただし、フレーズ指令が前の2つのアクセント句に接続する場合には、フレーズ指令の挿入を行なわない。

【0029】<ルールA-5>

修飾タイプが“並列”もしくは“並列補助”で、係り先距離が1の場合

累積のフレーズ成分をリセットし、大きさp h 2のフレーズ指令を挿入する。固有角周波数は $\alpha 1$ とする。

【0030】<ルールA-6>係り先距離が1より大きく、後接するアクセント句との間にポーズが存在しない場合は、後接するアクセント句のフレーズ指令をリセットし、大きさp h 2のフレーズ指令を挿入する。また、固有角周波数は $\alpha 1$ とする。

【0031】<ルールA-7>前節するアクセント句がポーズ句の先頭もしくはリセット後のフレーズ指令を伴い、モーラ数が3以下で、アクセント型が1の場合であって、前接するアクセント句と当該アクセント句との間にポーズが存在せず、当該アクセント句のモーラ数が9以上の場合、当該アクセント句のフレーズ指令をリセットし、大きさp h 2のフレーズ指令を挿入し、固有角周波数は $\alpha 1$ とする。また、前接するアクセント句のフレーズ指令の固有角周波数を $\alpha 3$ とする。一方、前接するアクセント句がポーズ句の先頭もしくはリセット後のフレーズ指令を伴い、モーラ数が3以下で、アクセント型が1の場合であって、前接するアクセント句と当該アクセント句の間にポーズが存在する場合は、前接するアクセント句のフレーズ指令の固有角周波数を $\alpha 3$ とする。

【0032】(3-2)ルールB

フレーズ指令が存在するアクセント句について、条件を満足するアクセント句にフレーズ指令を挿入する。

【0033】ルールBを適用する処理の流れは、以下の通りである。

- ・ステップ0: 文頭のアクセント句に着目する。
- ・ステップ1: 次のフレーズ指令が立つ直前のアクセント句(このアクセント句を以下A P Hと呼ぶことにする)までのモーラ数の合計を計算する。
- ・ステップ2: モーラ数の合計が14以下の場合で、次のフレーズ指令がない場合は終了し、次のフレーズ指令が存在する場合は、次のフレーズ指令が立つアクセント句に着目し、ステップ1に戻る。また、モーラ数の合計が14より大きい場合は、当該アクセント句からアクセント句A P Hの間で、係り先距離が最大であるアクセント句を選択する。該当するアクセント句が複数存在する場合は、当該アクセント句と前接するアクセント句との境界において、左右のモーラ数の和の差の絶対値が最小となるアクセント句を選択する。さらに該当するアクセント句が複数存在する場合は、最左アクセント句を選択

する。

- ・ステップ3：選択したアクセント句に大きさ ph_3 のフレーズ指令を挿入する。固有角周波数は $\alpha/2$ とする。
- ・ステップ4：ステップ1に戻る。

【0034】(3-3) ルールC

ルールCを適用する処理の流れは、以下の通りである。

- ・ステップ0：文頭のアクセント句に着目する。
- ・ステップ1：着目しているアクセント句に対して、以下に記述するルールC-1からルールC-3の条件を満足するルールを適用する。条件判定はルールの記述順に行なう。
- ・ステップ2：後節するアクセント句が存在する場合は、後接するアクセント句に着目し、ステップ1に戻る。後接するアクセント句が存在しない場合は、終了する。

【0035】以下にルールCに含まれる各ルールについて記述する。

<ルールC-1>文末のポーズ句が1つのアクセント句で構成されており、モーラ数が9以下の場合は、フレーズ指令の大きさを ph_1 とする。

<ルールC-2>ポーズ句の先頭でなく、ポーズ句の先頭から当該アクセント句までの間にフレーズ指令が重畳されておらず、当該アクセント句のアクセント型が2以上で、モーラ数が4以上の場合は、有核型のアクセント指令の大きさを acc_d2 とする。

<ルールC-3>ポーズ句の先頭から2番目のアクセント句において、リセットを伴わないフレーズ指令が存在し、モーラ数が5以下で、ポーズ句の先頭のアクセント句のフレーズ指令の大きさが ph_2 の場合は、フレーズ指令の大きさを ph_1 とする。また、この場合で、更に有核型のアクセント指令の大きさが acc_d1 の場合は、有核型のアクセント指令の大きさを acc_d2 とする。

【0036】以上のようにして、ステップS301による指令パラメータの生成を終えると、処理は、ステップS302に移る。ステップS302以降では、上記規則により生成された各パラメータの値の最適化を行なう。ここで、説明のため ph_1 , ph_2 , ..., α_3 などのパラメータを X_i ($i=1\sim N$) と表わし、パラメータ X_i の話者による変動を十分にカバーしうる範囲の変動幅を $\min_i \sim \max_i$ とし、その変動幅内でのステップ幅を $St * \ln F_0(t) = \ln F_{min}$

$$+ \sum_{i=1}^I A_{pi} \{G_{pi}(t - T_{0i})\} + \sum_{j=1}^J A_{aj} \{G_{aj}(t - T_{1j}) - G_{aj}(t - T_{2j})\} \quad (1)$$

【0044】ただし、 F_{min} は、声帯振動が可能な最低周波数である。ここで、 $G_{pi}(t)$, $G_{aj}(t)$ は、フレーズ

* ep_i とする。

【0037】まず、 $i=0$ から始めるために、ステップS302において $i=0$ をセットする。次に、ステップS303において、 $i \leq N$ をチェックし、 $i \leq N$ ならばステップS304に移る。そうでなければ、ステップS312に移る。

【0038】ステップS304では、 X_i に \min_i をセットする。そして、ステップS305において、 $X_i \leq \max_i$ ならば、ステップS307に移る。そうでなければ、ステップS306aへ進み、 X_i に X_{i_kouho} (暫定最小誤差を記録したときのパラメータ値のこと) をセットし、続いてステップS306bで i に $i+1$ をセットしてステップS303に戻る。

【0039】ステップS307では、変動中のパラメータ X_i を含む各パラメータのその時点での値を用いて、音声言語コーパス101中の各文の基本周波数パターンを改良型基本周波数生成モデルにしたがって計算し、ステップS308に移る。以下、改良型基本周波数パターン生成モデルについて説明する。

【0040】[改良型基本周波数パターン生成モデル] 改良型基本周波数パターン生成モデルは、上述した指令パラメータ生成規則により生成したパラメータをもとに基本周波数パターンを生成した場合の句末や文末が平坦化して聞こえるという欠点を解消したモデルである。

【0041】まず、「日本語単語アクセントの基本周波数パターンとその生成機構のモデル」(藤崎博也, 須藤寛, 日本音響学会誌, 27巻, 9号, pp.445~453, 1971年)において提案された基本周波数パターン生成過程モデルについて以下に説明する。

【0042】(1) 基本周波数パターン生成過程モデル 図5は、一般的な基本周波数パターン生成仮定モデルを説明する図である。文音声の基本周波数パターンは、一般的に複数個のフレーズ成分とアクセント成分とからなり、文中の i 番目のフレーズ指令の位置、大きさをそれぞれ T_{0i} , A_{pi} , j 番目のアクセント指令の始点、終点、大きさをそれぞれ T_{1j} , T_{2j} , A_{aj} とし、1文中でのフレーズ指令、アクセント指令の数をそれぞれ I , J とすると、時刻 t の関数としての基本周波数パターン $F_0(t)$ は、次式(1)で与えられる。

【0043】

【数1】

制御機構のインパルス応答関数、アクセント制御機構のステップ応答関数であり、 $t \geq 0$ の範囲では以下の式(2)、(3)で表される。

$$G_{pi}(t) = \alpha_i^2 t \exp(-\alpha t) \quad (2)$$

$$G_{aj}(t) = \min[1 - (1 + \beta_j t) \exp(-\beta_j t), \theta] \quad (3)$$

【0046】であり、 $t < 0$ の範囲では、以下の式(4)で表される。

$$G_{pi}(t) = G_{aj}(t) = 0$$

*【0045】
【数2】

*

※【0047】

※10 【数3】

(4)

【0048】なお、式(3)の記号 $\min[x, y]$ は、 x, y のうち小さい方をとることを意味しており、実際の音声でアクセント成分が有限の時間で上限に達することに対応し、 $\theta \leq 1$ である。 α_i, β_j は、それぞれフレーズ制御機構およびアクセント制御機構の固有角周波数である。

【0049】(2)改良型基本周波数パターン生成過程モデル

ポーズ句の先頭のフレーズ指令、もしくは、リセットを伴うフレーズ指令から生成されるフレーズ成分について以下の補正を行なう。ただし、以下の説明において、フレーズ指令の有効範囲とは、フレーズ指令の存在するアクセント句の先頭モーラの開始時点から、次のポーズの直前、もしくは、次のリセットを伴うフレーズ指令の直前のアクセント句の最終モーラの終了時点までを指す。

【0050】①フレーズ指令の有効範囲が1000 msec以下の場合

フレーズ指令の有効範囲が1000 msec以下の場合、フレーズ指令の有効範囲すべてに対して、基本周波数パターンに負の傾き k_1 をもつ直線を重畳する。なお、直線の傾きを決める係数 k_1 は、話者により異なるものとする。また、傾き k_1 を有する直線は図6の点Aを通るものである。

【0051】②フレーズ指令の有効範囲が1000 msecより長い場合

フレーズ指令の有効範囲に含まれる最後のアクセント句に対して、基本周波数パターンに負の傾き k_2 をもつ直線を重畳する。ただし、直線の傾きを決める係数 k_2 は、話者により異なるものとする。また、傾き k_2 を有する直線は図7の点Bを通るものである。

【0052】図6はフレーズ指令の有効範囲が1000 msec以下の場合の基本周波数の改良方法を概念的に示した図である。図7は、フレーズ指令の有効範囲が1000 msecを越える場合の基本周波数の改良方法を概念的に示した図である。重畳する直線は、フレーズ指令の有効範囲が1000 msec以下の場合、図6に示されるように点Aを通る傾き k_1 の直線であり、当該フレーズ指令有効範囲の全体に重畳される。また、フレーズ

ズ指令の有効範囲が1000 msec以上の場合、図7に示されるように点Bを通る傾き k_2 の直線であり、当該フレーズ指令有効範囲の末尾のアクセント句の範囲に重畳される。

【0053】直線により傾斜をつけた場合、末尾に近いアクセント句ほど基本周波数が低くなる傾向が生じる。このため、フレーズ指令の有効範囲の違い(本例では、1000 msecより長いのか短いのか)で場合分けし、有効範囲が長い場合に、基本周波数が低くなり過ぎるのを防止する。なお、図6、図7のいずれにおいても、補正直線の重畳後も、フレーズ指令があるアクセント句の先頭モーラの開始時点が元の基本周波数のままとする。

【0054】再び図4において、ステップS308では、実発声から求めた基本周波数パターンと上記アルゴリズムより求めた基本周波数パターンとの対数スケール上での最小二乗誤差を計算して、ステップS306に移る。

【0055】ステップS306では、ステップS305で求めた最小二乗誤差がそれまでに求めた最小二乗誤差の最小値(以下、暫定最小誤差)より小さいかどうかを判別する。小さくない場合は、ステップS311へ進み、 $X_i = X_i + \text{Step}_i$ として、ステップS305に戻る。一方、ステップS309において、暫定最小誤差よりも小さい場合はステップS310に移り、最小誤差となった誤差値とそのときのパラメータの値を保存する。そして、ステップS311で $X_i = X_i + \text{Step}_i$ として、ステップS305に戻る。

【0056】また、ステップS312では、ステップS303からステップS311までの過程を通る前と後で誤差最小値が更新されたかどうかを判別する。更新された場合は、ステップS304へ処理を分岐し、 $i = 0$ として、ステップS305に移る。一方、更新されなかった場合には、本処理を終了する。

【0057】次に図8に示すフローチャートを参照して、オンライン処理における動作を説明する。

【0058】音声言語コーパスの替わりに、入力文に対する言語解析の結果情報を入力とし(ステップS401)、指令パラメータ生成規則102と同じ指令パラメータ生成規則111を使って指令パラメータを生成し、

パラメータ保持部113に保持する（ステップS402）。

【0059】ステップS403では、パラメータ保持部113に保持している指令パラメータに基づいて、話者別のパラメータ値を保持した話者別パラメータ値保持部114からパラメータ値を得る。そして、ステップS404にて、ステップS403で得たパラメータ値を用いて基本周波数ボタンを生成する。基本周波数ボタンの生成は、オフライン処理に準ずる。生成された基本周波数ボタンは基本周波数ボタン保持部116に保持される。最後にステップS405にて、生成された基本周波数ボタンを用いて音声合成を行い、本処理を終了する。

【0060】以上説明したように、上記実施形態によれば、話者の特徴が反映され、かつ、基本周波数の平坦化が防止された自然な合成音声を生成することが可能な基本周波数ボタンを求めることができる。

【0061】なお、上述した実施形態に関しては、各種の変形が可能であり、その例を示すと次の通りである。

【0062】（1）上記実施形態では、誤差計算を対数スケール上での基本周波数に対して最小二乗法を施す場合について説明したが、これに限定されるものでない。例えば、線形（リニア）スケール上での基本周波数に対して最小二乗法を施すようにしても良い。

【0063】（2）上記実施形態では、指令パラメータ生成規則において補正直線の傾きを決定するためのパラメータを、フレーズ指令の有効期間が1000msec以下の場合とそれを越える場合とで分けたが、これに限定されるものではない。例えば、分けなくてもよい。もちろん、1000msecというのも一例であり、話者ごとに決定されるパラメータとしてもよい。

【0064】（3）上記実施形態では、指令パラメータ生成規則の具体的な例を挙げたが、これに限定されるものではなく、指令パラメータ生成規則を他の規則に取り替えてもよい。

【0065】（4）上記実施形態では、オフライン処理においてパラメータの値を求める場合の各パラメータの初期値として、変動幅の最小値を与えるが、これに限定されるものではない。たとえば、初期値は変動幅の最大値や平均値、または、コーパスから得られる平均値を設定するようにしても良い。

【0066】（5）上記実施形態では、オフライン処理においてパラメータの値を求める場合の各パラメータの初期値として、変動幅の最小値を与える場合について説明したが、これに限定されるものではなく、パラメータの初期値は変動幅内で発生させた乱数を与えてもよいものとする。

【0067】（6）上記実施形態では、ステップS308で誤差の最小値が更新されなかった場合に処理を終了しているが、これに限定されるものではない。たとえば、パラメータの初期値を変化させることにより得られ

る結果が異なるので、初期値を乱数で与える試行を十分な回数（100回など）行ない、そのなかで最も誤差が最小になるパラメータの値の組をパラメータセットとするようにしてもよい。或いは、パラメータを変動させる順番を代えることにより得られる結果が異なるので、パラメータの変動順序を乱数により決定する試行を十分な回数（100回など）行ない、そのなかで最も誤差が最小になるパラメータの値の組をパラメータセットとするようにしても良い。

【0068】（7）上記実施形態では、変動幅を固定して試行する場合について記述したが、これに限定されるものでなく、十分な試行回数のうち得られたパラメータセットにおいて、変動幅を小さくして、より局所的な範囲において、より精度よくパラメータの値を求めることを繰り返すようにしてもよい。例えば、変動幅10～20、ステップ幅1で試行した結果、パラメータの値14が最適値と求まった場合、変動幅を最適値の中心にステップ幅×2（＝12～16）で自動的に設定し、ステップ幅をもとのステップ幅の1/5（＝0.2）に自動的に設定し、より最適なパラメータ値を求めるようにしても良い。

【0069】（8）上記実施形態においては、指令パラメータ生成規則102や音声言語コーパス101や話者別パラメータ値辞書114をディスク上に実現する場合について説明したが、これに限定されるものではなく、メモリ（RAM）など任意の記憶媒体を用いて実現してもよい。

【0070】（9）上記実施形態においては、各部を同一の計算機上で構成する場合について説明したが、これに限定させるものではなく、ネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

【0071】（10）上記実施形態においては、プログラムを制御メモリ（ROM）に保持する場合について説明したが、これに限定されるものではなく、任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

【0072】なお、本発明は、複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に適用してもよい。前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記録媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【0073】この場合、記録媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。

【0074】プログラムコードを供給するための記録媒

体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

【0075】また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOSなどが実際の処理の一部または全部を行ない、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0076】更に、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0077】

【発明の効果】以上説明したように、本発明によれば、話者に応じた基本周波数パターンを生成可能とし、自然な

合成音声を得ることが可能となる。

【0078】

【図面の簡単な説明】

【図1】本発明の一実施形態に係る音声合成装置のオフライン処理における機能構成を示すブロック図である。

【図2】本発明の一実施形態に係る音声合成装置のオンライン処理における機能構成を示すブロック図である。

【図3】本実施形態による音声合成装置の構成を示すブロック図である。

【図4】本実施形態の音声合成装置によるオフライン処理の動作を説明するフローチャートである。

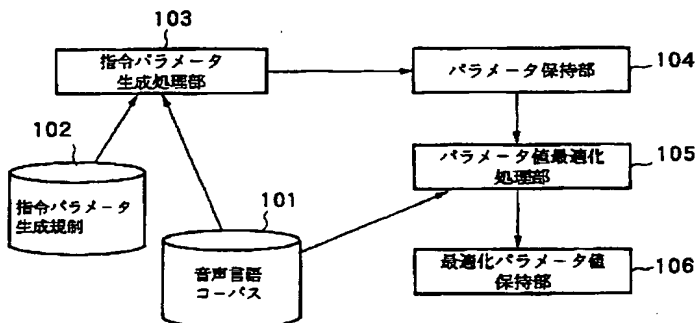
【図5】一般的な基本周波数パターン生成仮定モデルを説明する図である。

【図6】フレーズ指令の有効範囲が1000msec以下の場合の基本周波数の改良方法を概念的に示した図である。

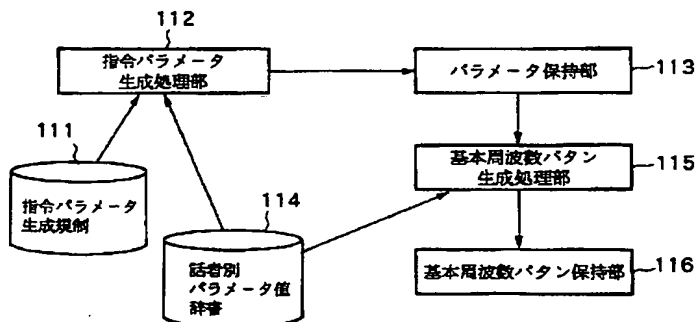
【図7】フレーズ指令の有効範囲が1000msecを超える場合の基本周波数の改良方法を概念的に示した図である。

【図8】本実施形態の音声合成装置によるオンライン処理の動作を説明するフローチャートである。

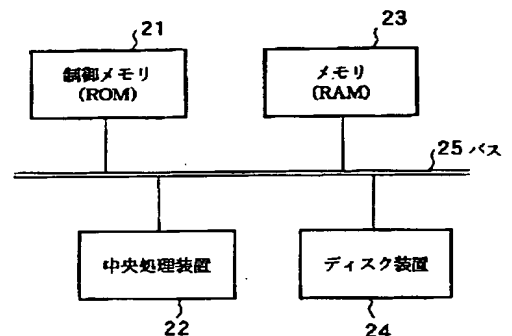
【図1】



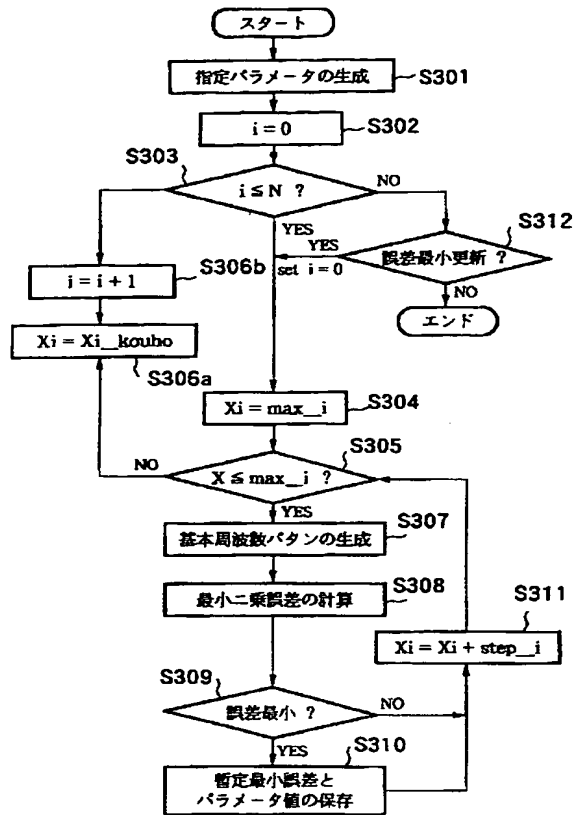
【図2】



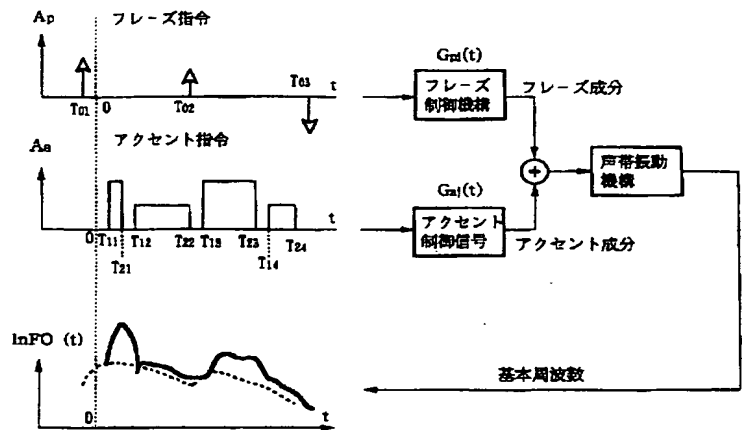
【図3】



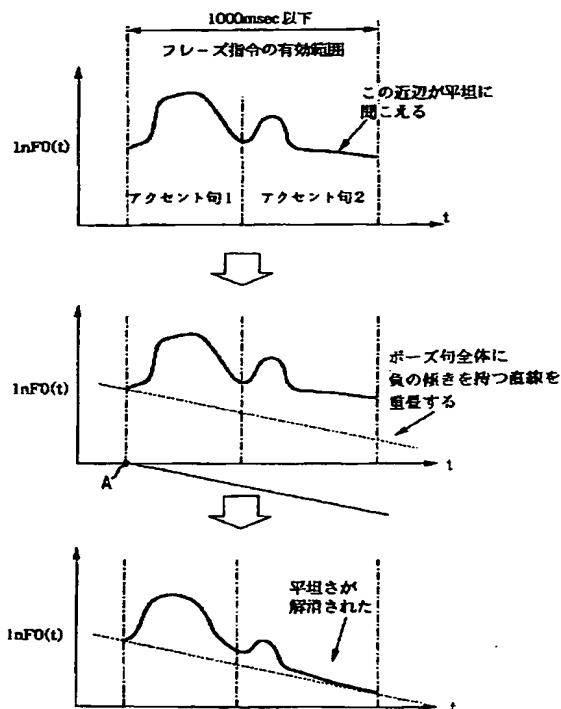
【図4】



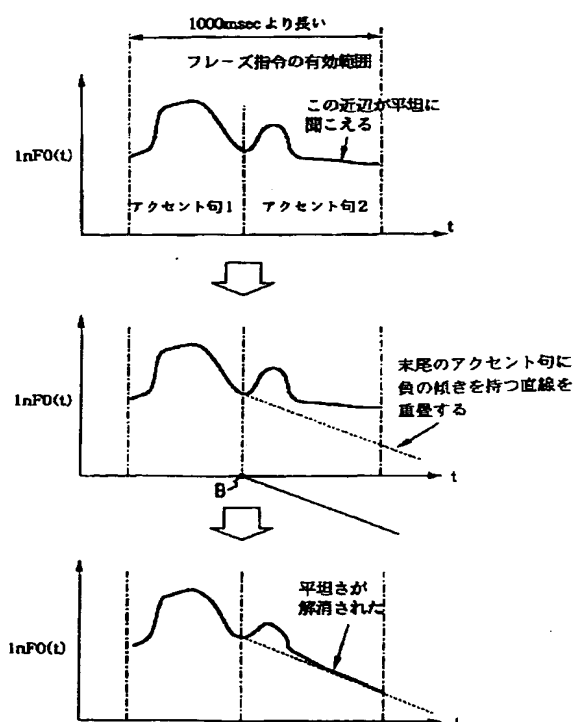
【図5】



【図6】



【図7】



【図8】

